Nonparametric adaptive time-dependent multivariate function estimation

*Jérémie Bigot**, DMIA/ISAE - University of Toulouse
Theofanis Sapatinas†, University of Cyprus

November 1, 2012

Abstract

We consider the nonparametric estimation problem of time-dependent multivariate functions observed in a presence of additive cylindrical Gaussian white noise of a small intensity. We derive minimax lower bounds for the L^2-risk in the proposed spatio-temporal model as the intensity goes to zero, when the underlying unknown response function is assumed to belong to a ball of appropriately constructed inhomogeneous time-dependent multivariate functions, motivated by practical applications. Furthermore, we propose both non-adaptive linear and adaptive non-linear wavelet estimators that are asymptotically optimal (in the minimax sense) in a wide range of the so-constructed balls of inhomogeneous time-dependent multivariate functions. The usefulness of the suggested adaptive nonlinear wavelet estimator is illustrated with the help of simulated and real-data examples.

Keywords: Adaptivity; Besov spaces; Block Thresholding; Minimax Estimators; Time-dependent image processing, Wavelet Analysis.
AMS classifications: Primary 62G05, 62G08, 62G20; Secondary 62H35.

1 Introduction

The nonparametric estimation problem of high-dimensional objects has been considered in the literature over the last three decades. With the help of appropriate balls in function spaces, such as, Hölder, Sobolev or Besov balls, that measure smoothness of the unknown underlying high-dimensional object, asymptotical (as the sample size goes to infinity) optimal properties (in the minimax sense) of various linear and non-linear estimators, such as, kernel, spline or wavelet estimators, have been obtained (see, e.g., [Wahba, 1990], [Korostelëv and Tsybakov, 1993] (regression setting) and [Klemelä, 2009] (density setting), and the references therein).

These optimality properties were studied by [Chow et al., 2001] in the case of time-dependent multivariate response functions. By following a trend to derive theoretical properties,

*Institut Supérieur de l'Aéronautique et de l'Espace, Département Mathématiques, Informatique, Automatique, 10 Avenue Édouard-Belin, BP 54032-31055, Toulouse CEDEX 4, France. Email: jeremie.bigot@isae.fr

†Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678, Nicosia, Cyprus. Email: fanis@ucy.ac.cy

[Chow et al., 2001] considered a "continuous-time" model for the estimation problem of time-dependent multivariate functions observed in a presence of additive cylindrical Gaussian white noise, that is, they considered

$$dY_\epsilon(t, \underline{x}) = \boldsymbol{f}(t, \underline{x})d\underline{x} + \epsilon dW(t, \underline{x}), \tag{1.1}$$

where $t \in T$ (T is a compact subset of $\mathbb{R}$) is the time variable, $\underline{x} \in \mathcal{X}$ ($\mathcal{X}$ is a compact subset of $\mathbb{R}^d$, $d \geq 1$) is the space variable, $\boldsymbol{f} \in \mathbb{L}^2(T \times \mathcal{X})$ is the time-dependent multivariate function that we wish to estimate, $dW(t, \underline{x})$ is a cylindrical orthogonal Gaussian random measure (representing additive noise in the measurements), and $\epsilon > 0$ is a small level of noise, that may let be going to zero for studying asymptotic properties.

A formal definition of a cylindrical orthogonal Gaussian random measure can be found in Section 2.1 of [Chow et al., 2001]. Moreover, we understand (1.1) in a generalized sense, that is, the observable elements are treated as linear functionals, so that the process $Y_\epsilon(t, \underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$, is correctly defined (see Section 6.1). Also, without loss of generality, in the sequel, we assume that $T = [0, 1]$ and $\mathcal{X} = [0, 1]^d$.

Assume periodic assumptions in each argument of $\boldsymbol{f}(t, \underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$. Consider Hölder continuity in $L^2(\mathcal{X})$ on the derivatives of $\boldsymbol{f}(t, \underline{x})$ with respect to $t \in T$, uniformly over $\underline{x} \in \mathcal{X}$, and Hölder continuity in $L^2(T)$ on the partial derivatives of $\boldsymbol{f}(t, \underline{x})$ with respect to the elements of $\underline{x} \in \mathcal{X}$, uniformly over $t \in T$. Then, under known a-priori smoothness (i.e., knowing the involved Hölder parameters) of $\boldsymbol{f}(t, \underline{x})$, [Chow et al., 2001] constructed a non-adaptive kernel-projection (linear) estimator and obtained an asymptotical (as $\epsilon \to 0$) upper bound of its L^2-risk (on $\mathcal{X}$), uniformly over a set $T_1 \subset T$, that depends on ϵ and the involved smoothness parameters (see, [Chow et al., 2001], Theorem 4.1). Moreover, they have showed that, asymptotically, this upper bound cannot be improved (see [Chow et al., 2001], Lemma 5.3), thus establishing the asymptotical optimality (in the minimax sense) of their suggested estimator.

Our aim is twofold. From a theoretical point of view, we extend the asymptotical optimal convergence rates derived in [Chow et al., 2001]. In particular, when smoothness is measured in appropriate balls of inhomogeneous functions, constructed with the help of tensor-product wavelet bases and Besov spaces, with or without a-priori knowledge of the involved smoothness parameters, we construct, respectively, non-adaptive linear (projection) or adaptive non-linear (block-thresholding) wavelet estimators that achieve the established asymptotical optimal convergence rates under the L^2-risk. From a practical point of view, we demonstrate the usefulness of the suggested adaptive nonlinear wavelet thresholding estimator in practical applications. In particular, we show the superiority of the suggested estimator in terms of average mean squared error over pixel by pixel and slice by slice wavelet denoising estimators, both with universal thresholds.

The paper is organized as follows. Section 2 provides a motivating example. Section 3 contains a brief summary of the tensor-product wavelet bases and standard Besov spaces while Section 4 discusses the function spaces that we consider to appropriately model the considered inhomogeneous time-dependent multivariate functions. Section 5 contains the minimax lower bounds for the L^2-risk. Section 6 introduces both non-adaptive linear and adaptive non-linear wavelet estimators and provides their minimax upper bounds for the L^2-risk in a wide range of the so-constructed balls of inhomogeneous time-dependent multivariate functions. Section

7 demonstrates the usefulness of the suggested adaptive nonlinear wavelet estimator with the help of simulated and real-data examples. Section 8 contains some concluding remarks. Finally, Section 9 (Appendix) provides two technical lemmas that are used in the proofs of the main theoretical results.

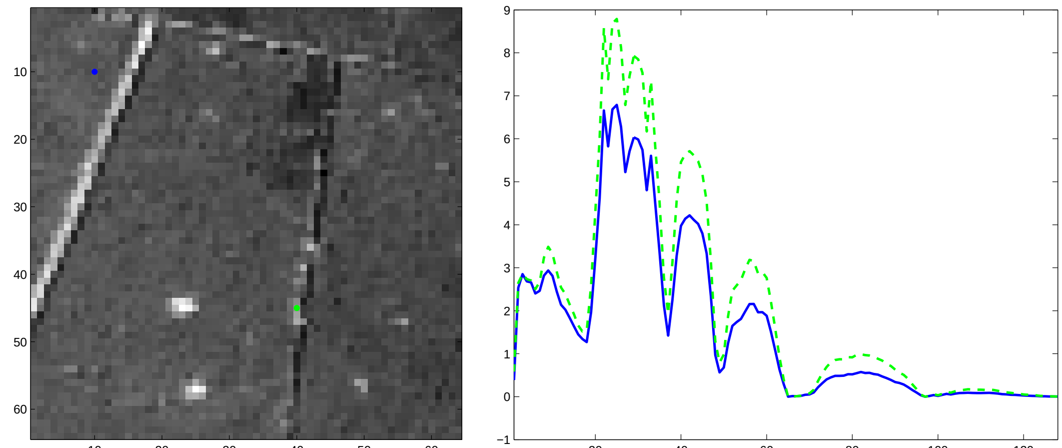
2 A motivating example

Increasingly, scientific studies yield time-dependent d-dimensional images, in which the observed data consist of sets of curves recorded on the pixels of d-dimensional images observed at different times or wavelengths, see e.g. [Antoniadis et al., 2009]. Examples include temporal brain response intensities measured by functional magnetic resonance imaging (fMRI) [Whitcher et al., 2005], satellite remote sensing images of landscapes [Ju et al., 2005], and functional brain mapping using electroencephalography (EEG) and magnetoencephalography (MEG) [Ou et al., 2009]. In many applications, the measured curves tend to be spiky and this requires flexible adaptive and local modeling of their variations. The high dimensionality and noise that characterize such time-dependent images makes difficult the estimation of the evolution of each pixel intensity over time (or wavelength).

We now discuss a specific application that motivates the estimation of $\boldsymbol{f}$ in model (1.1), and the choice of the function spaces that we use to measure the smoothness of $\boldsymbol{f}$ (see Section 4). An example of application and data fitting into model (1.1) is satellite remote sensing imaging of landscapes, where the data are in the form of a multiband satellite 2-dimensional image of remote sensing measurements in various spectral bands of an area that contains roads, forests, vegetation, lakes and fields, see [Antoniadis et al., 2009]. As an illustrative example, we display in Figure 2.1(a) a typical temporal (or wavelength) slice (i.e., 2-dimensional grey-level image), and we plot in Figure 2.1(b) two curves (1-dimensional signals) corresponding to two selected pixels highlighted in blue and green in Figure 2.1(a). With respect to model (1.1), Figure 2.1(a) corresponds to a noisy version of $\underline{x} \mapsto \boldsymbol{f}(t, \underline{x})$ for some fixed $t \in T$, while Figure 2.1(b) corresponds to a noisy version of $t \mapsto \boldsymbol{f}(t, \underline{x})$ for some fixed $\underline{x} \in \mathcal{X}$.

3 Wavelets and Besov spaces

We briefly consider tensor-product wavelet bases of $\mathbb{L}^2(\mathbb{R}^d)$, $d \geq 1$, and recall some of their properties; for a detailed description of their construction, we refer to [Mallat, 2009]. Assume that we have at our disposal a 1-dimensional scaling function (i.e., a father wavelet) ϕ and a 1-dimensional wavelet function (i.e., a mother wavelet) ψ, both with compact supports. The scaling and wavelet functions of ϕ and ψ, at scale j (i.e., at resolution level 2^j) will be denoted by ϕ_λ and ψ_λ, respectively, where the index λ summarizes both the usual scale and space parameters j and k. In other words, for $d = 1$, we set $\lambda = (j, k)$ and denote $\phi_{j,k}(\cdot) = 2^{j/2}\phi(2^j \cdot -k))$ and $\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j \cdot -k))$. For $d \geq 2$, the notation ψ_λ stands for the adaptation of scaling and wavelet functions to $\mathbb{R}^d$ (see [Mallat, 2009], Chapter 7). The notation $|\lambda| = j$ will be used to denote a wavelet at scale j, while $|\lambda| < j$ denotes a wavelet at scale j', with $j_0 \leq j' < j$, where j_0 denotes the coarse level of approximation (usually called the primary resolution level). With the above notation, we assume that



(a) (b)

Figure 2.1: Satellite remote sensing image (64×64 pixels, over 128 wavelengths): (a) a 2-dimensional image measured at a specific wavelength; (b) evolution over wavelength of the intensities of the two pixels in green and blue marked in the image shown in (a).

- the scaling functions $(\phi_\lambda)_{|\lambda|=j}$ span a finite dimensional space V_j within a multiresolution hierarchy $V_0 \subset V_1 \subset \ldots \subset L^2(\mathbb{R}^d, dy)$, such that $\dim(V_j) = 2^{jd}$.

- the scaling functions $(\phi_\lambda)_{|\lambda|=j}$ form an orthonormal basis of V_j and the wavelets $(\psi_\lambda)_{|\lambda|=j}$ form an orthonormal basis of W_j (with W_j being the orthogonal complement of V_j into V_{j+1}).

- Let $\mathcal{Y}$ be a compact subset of $\mathbb{R}^d$, $d \geq 1$. Assuming periodicity in each argument of $\underline{y} \in \mathcal{Y}$, and using standard wavelet bases ($d = 1$) or tensor-product wavelet bases ($d \geq 2$) of $L^2(\mathcal{Y})$ (see, e.g. [Mallat, 2009], Chapter 7), any $f \in \mathbb{L}^2(\mathcal{Y})$ can be decomposed as

$$f(\underline{y}) = \sum_{|\lambda|=j_0} c_\lambda \phi_\lambda(\underline{y}) + \sum_{j=j_0}^{+\infty} \sum_{|\lambda|=j} \beta_\lambda \psi_\lambda(\underline{y}), \quad \underline{y} \in \mathcal{Y},$$

where

$$c_\lambda = \langle f, \phi_\lambda \rangle_{\mathcal{Y}} \quad \text{and} \quad \beta_\lambda = \langle f, \psi_\lambda \rangle_{\mathcal{Y}}.$$

In order to simplify the notation, as it is commonly used, we write $(\psi_\lambda)_{|\lambda|=j_0-1}$ for $(\phi_\lambda)_{|\lambda|=j_0}$, and, thus, f can be written in the compact form

$$f(\underline{y}) = \sum_{j=j_0-1}^{+\infty} \sum_{|\lambda|=j} \alpha_\lambda \psi_\lambda(\underline{y}), \quad \underline{y} \in \mathcal{Y},$$

where α_λ denotes either the scaling coefficients c_λ or the wavelet coefficients β_λ.

Consider also the following balls of (inhomogeneous) Besov spaces.

Let $s_1 > 0$ be a smoothness parameter in the domain $\mathcal{Y}_1$ (that is, $\mathcal{Y}$ with $d \geq 2$), and let $1 \leq p_1, q_1 \leq +\infty$. Let $(\psi_\lambda)_{|\lambda|=j}$, $j \geq j_0$, be the (periodic) d-dimensional (tensor-product)

compactly supported orthonormal wavelet basis of $L^2(\mathcal{Y}_1)$, with the convention that $(\psi_\lambda)_{|\lambda|=j_0-1}$ denotes the scaling functions $(\phi_\lambda)_{|\lambda|=j_0}$. Assume that the 1-dimensional scaling function ϕ and the 1-dimensional wavelet function ψ are τ_1-times continuously differentiable (regularity of the wavelet system (ϕ,ψ)) with $0 < s_1 < \tau_1$, and assume that $s_1 + d(1/2 - 1/p_1) > 0$. Define the norm $\|\cdot\|^{s_1}_{p_1,q_1}$ by

$$\|f\|^{s_1}_{p_1,q_1} = \left(\sum_{j=j_0-1}^{+\infty} 2^{j(s_1+d(1/2-1/p_1))q_1} \left(\sum_{|\lambda|=j} |\langle f, \psi_\lambda\rangle|^{p_1} \right)^{q_1/p_1} \right)^{1/q_1},$$

with the respective above sums replace by maximum if $p_1 = +\infty$ and/or $q_1 = +\infty$. Then, the norm $\|\cdot\|^{s_1}_{p_1,q_1}$ is equivalent to the traditional Besov norm (see e.g [Härdle et al., 1998] for further details), and one can thus define the following Besov ball of radius $A_1 > 0$

$$B^{s_1}_{p_1,q_1}(A_1) = \left\{ f \in \mathbb{L}^2(\mathcal{Y}_1),\ \|f\|^{s_1}_{p_1,q_1} \leq A_1 \right\}.$$

Let $s_2 > 0$ be a smoothness parameter in the domain $\mathcal{Y}_2$ (that is $\mathcal{Y}$ with $d = 1$), and let $1 \leq p_2, q_2 \leq +\infty$. Let $(\tilde{\psi}_{m,\ell})_{m=m_0-1,\ell=0,\ldots,2^m-1}$ be a (periodic) 1-dimensional compactly supported orthonormal wavelet basis of $\mathbb{L}^2(\mathcal{Y}_2)$, with the convention that $(\tilde{\psi}_{m_0-1,\ell})_{\ell=0,\ldots,2^{m_0}-1}$ denotes the scaling functions $(\tilde{\phi}_{m_0,\ell})_{\ell=0,\ldots,2^{m_0}-1}$, where m_0 is the coarse (primary) resolution level. Assume that the corresponding 1-dimensional scaling function $\tilde{\phi}$ and the 1-dimensional wavelet function $\tilde{\psi}$ are τ_2-times continuously differentiable (regularity of the wavelet system $(\tilde{\phi},\tilde{\psi})$) with $0 < s_2 < \tau_2$, and assume that $s_2 + 1/2 - 1/p_2 > 0$. Define the norm $\|\cdot\|^{s_2}_{p_2,q_2}$ by

$$\|g\|^{s_2}_{p_2,q_2} = \left(\sum_{m=m_0-1}^{+\infty} 2^{m(s_2+1/2-1/p_2)q_2} \left(\sum_{m=0}^{2^m-1} |\langle g, \tilde{\psi}_{m,\ell}\rangle|^{p_2} \right)^{q_2/p_2} \right)^{1/q_2},$$

with the respective above sums replace by maximum if $p_2 = +\infty$ and/or $q_2 = +\infty$. Then, as noticed above, one can define the following Besov ball of radius $A_2 > 0$

$$B^{s_2}_{p_2,q_2}(A_2) = \left\{ g \in \mathbb{L}^2(\mathcal{Y}_2),\ \|g\|^{s_2}_{p_2,q_2} \leq A_2 \right\}.$$

4 Smoothness assumptions on the time-dependent multivariate response function

The statistical problem that we consider below is the estimation of the unknown time-dependent multivariate response function $\boldsymbol{f}(t,\underline{x})$, $\underline{x} \in \mathcal{X}$, $t \in T$, based on observations from model (1.1). Motivated by the practical application discussed in Section 2, in order to derive the asymptotical (as $\epsilon \to 0$) optimal (in the minimax sense) rates of convergence (for the L^2-risk), we consider the following functional space to model $\boldsymbol{f}(t,\underline{x})$, $\underline{x} \in \mathcal{X}$, $t \in T$.

First, let us assume that, for each $t \in T$, the mapping $\underline{x} \mapsto \boldsymbol{f}(t,\underline{x})$ belongs to $\mathbb{L}^2(\mathcal{X})$. Let $\Lambda = \{\lambda, |\lambda| = j\}_{j_0-1 \leq j \leq +\infty}$. For each $t \in T$, the (periodic) d-dimensional wavelet basis $(\psi_\lambda)_{\lambda\in\Lambda}$ is used to decompose $\boldsymbol{f}(t,\underline{x})$ as

$$\boldsymbol{f}(t,\underline{x}) = \sum_{j=j_0-1}^{+\infty} \sum_{|\lambda|=j} \boldsymbol{\alpha}_\lambda(t)\psi_\lambda(\underline{x}) \quad \text{with} \quad \boldsymbol{\alpha}_\lambda(t) = \langle \boldsymbol{f}(t,\cdot), \psi_\lambda\rangle_{\mathbb{L}^2(\mathcal{X})}, \quad \underline{x} \in \mathcal{X}. \tag{4.1}$$

Then, for each $\lambda \in \Lambda$, we assume that the mapping $t \mapsto \boldsymbol{\alpha}_\lambda(t)$ belongs to $\mathbb{L}^2(T)$. For each $\lambda \in \Lambda$, the (periodic) 1-dimensional wavelet basis $(\tilde{\psi}_{m,\ell})_{m=m_0-1,\ell=0,\ldots,2^m-1}$ is used to decompose $\boldsymbol{\alpha}_\lambda(t)$ as

$$\boldsymbol{\alpha}_\lambda(t) = \sum_{m=m_0-1}^{+\infty} \sum_{\ell=0}^{2^m-1} \tilde{\alpha}_{\lambda,m,\ell} \tilde{\psi}_{m,\ell}(t) \quad \text{with} \quad \tilde{\alpha}_{\lambda,m,\ell} = \langle \boldsymbol{\alpha}_\lambda, \tilde{\psi}_{m,\ell} \rangle_{\mathbb{L}^2(T)}, \quad t \in T. \tag{4.2}$$

Finally, by assuming that the mapping $(t, \underline{x}) \mapsto \boldsymbol{f}(t, \underline{x})$ belongs to $\mathbb{L}^2(T \times \mathcal{X})$ for any $t \in T$ and $x \in \mathcal{X}$, and consider the corresponding tensor product wavelet basis, $\boldsymbol{f}(t, \underline{x})$ can thus be decomposed as

$$\boldsymbol{f}(t, \underline{x}) = \sum_{j=j_0-1}^{+\infty} \sum_{|\lambda|=j} \sum_{m=m_0-1}^{+\infty} \sum_{\ell=0}^{2^m-1} \tilde{\alpha}_{\lambda,m,\ell} \tilde{\psi}_{m,\ell}(t) \psi_\lambda(\underline{x}), \quad t \in T, \quad \underline{x} \in \mathcal{X}.$$

We are now ready to introduce the following definition in order to characterize the smoothness of the time-dependent multivariate function $\boldsymbol{f}(t, \underline{x})$, $t \in T$, $x \in \mathcal{X}$.

Definition 1. *Let $A_1 > 0$ and $A_2 > 0$ be constants. Let $s_1 > 0$ be a smoothness parameter in space domain $\mathcal{X}$ and $s_2 > 0$ be a smoothness parameter in time domain T, such that $0 < s_1 < \tau_1$ and $0 < s_2 < \tau_2$, where τ_1 and τ_2 are the regularity parameters of the wavelet systems (ϕ, ψ) and $(\tilde{\phi}, \tilde{\psi})$, respectively. Let $1 \le p_1, q_1 \le +\infty$, $1 \le p_2, q_2 \le +\infty$, and assume that $s_1 + d(1/2 - 1/p) > 0$ and $s_2 + 1/2 - 1/p_2 > 0$. Let $p = (p_1, p_2)$ and $q = (q_1, q_2)$. Define $\mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)$ as the following ball of functions in $\mathbb{L}^2(T \times \mathcal{X})$:*

$$\mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2) = \left\{ \boldsymbol{f} \in \mathbb{L}^2(T \times \mathcal{X}) \mid \sup_{t \in T} \{ \|\boldsymbol{f}(t, \cdot)\|^{s_1}_{p_1,q_1} \} \le A_1 \ \text{and} \ \|\boldsymbol{\alpha}_\lambda\|^{s_2}_{p_2,q_2} \le A_\lambda \ \text{for all} \ \lambda \in \Lambda \right\},$$

where, for each $t \in T$,

$$\|\boldsymbol{f}(t, \cdot)\|^{s_1}_{p_1,q_1} = \left(\sum_{j=j_0-1}^{+\infty} 2^{j(s_1 + d(1/2 - 1/p_1))q_1} \left(\sum_{|\lambda|=j} |\langle \boldsymbol{f}(t, \cdot), \psi_\lambda \rangle|^{p_1} \right)^{q_1/p_1} \right)^{1/q_1},$$

for each $\lambda \in \Lambda$,

$$\|\boldsymbol{\alpha}_\lambda\|^{s_2}_{p_2,q_2} = \left(\sum_{m=m_0-1}^{+\infty} 2^{m(s_2 + 1/2 - 1/p_2)q_2} \left(\sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^{p_2} \right)^{q_2/p_2} \right)^{1/q_2},$$

with

$$\tilde{\alpha}_{\lambda,m,\ell} = \int_{T \times \mathcal{X}} \boldsymbol{f}(t, \underline{x}) \tilde{\psi}_{m,\ell}(t) \psi_\lambda(\underline{x}) dt d\underline{x},$$

and $(A_\lambda)_{\lambda \in \Lambda}$ is a a set of positive constants such that

$$\sum_{j=j_0-1}^{+\infty} \sum_{|\lambda|=j} A_\lambda^2 \le A_2^2. \tag{4.3}$$

Assuming that $\boldsymbol{f}(t,\underline{x}) \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$ means that the multivariate function $\boldsymbol{f}(t,\cdot)$ belong to $B^{s_1}_{p_1,q_1}(A_1)$, uniformly over $t \in T$. This assumption also means that the smoothness of the wavelet coefficients $(\boldsymbol{\alpha}_\lambda(\cdot))_{\lambda\in\Lambda}$ over time $t \in T$ is measured by the parameter s_2 through a Besov ball $B^{s_2}_{p_2,q_2}(A_\lambda)$ whose radius satisfies equation (4.3). It implies that $\sup_{\lambda\in\Lambda}\{A_\lambda\} \le A_2$ and, more importantly, that $\lim_{j\to+\infty,\,|\lambda|=j} A_\lambda = 0$ so that the Besov norm $\|\boldsymbol{\alpha}_\lambda\|^{s_2}_{p_2,q_2}$ goes to zero as the resolution level of the time-dependent wavelet coefficients $\boldsymbol{\alpha}_\lambda(\cdot)$ goes to infinity. In practical applications, it correspond to the assumption that the high-resolution energy of a time-dependent multivariate function, when integrated over time, is going to zero. (In order to simplify the notation, we have dropped the dependence of $\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$ on $(A_\lambda)_{\lambda\in\Lambda}$.)

To motivate the definition of the functional space $\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$, let us consider the real-data example on satellite remote sensing data discussed in Section 2. For this time-dependent 2-dimensional image, we display in Figure 4.2 the curve $t \mapsto \boldsymbol{\alpha}_\lambda(t)$ for two types of wavelet coefficients, one at a low resolution level ($|\lambda| = 3$) and another one at the highest resolution level ($|\lambda| = 5$). Clearly, the curve at the highest resolution level has a smallest amplitude which is consistent with the decay of A_λ as $|\lambda|$ increases in the definition of $\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$. Moreover, due the shape of the curves in Figure 4.2, it seems reasonable to assume that the functions $\boldsymbol{\alpha}_\lambda(\cdot)$ have the same degree of smoothness s_2 across different resolution levels.

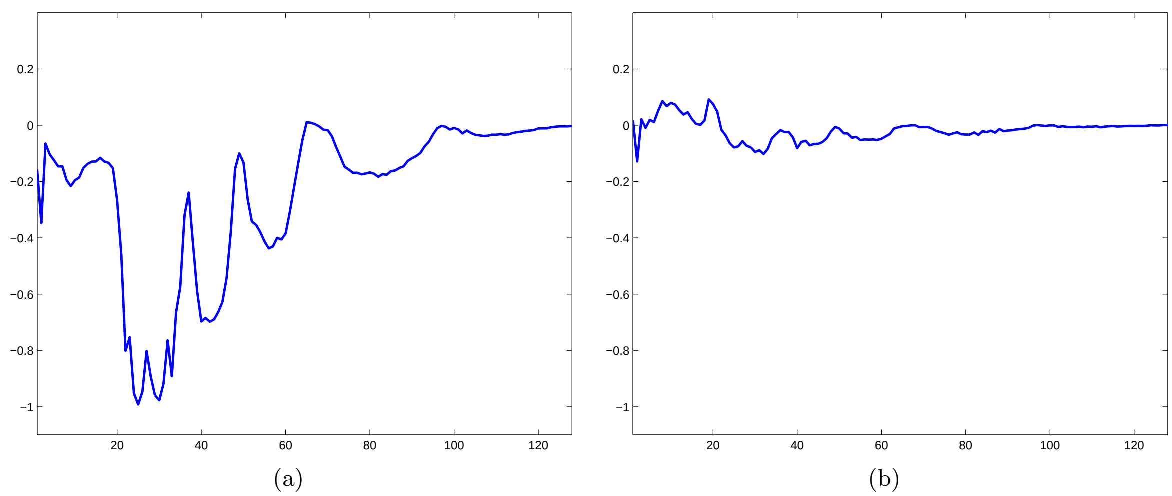


Figure 4.2: Satellite remote sensing image. Evolution of the curve $t \mapsto \boldsymbol{\alpha}_\lambda(t)$ for (a) a wavelet coefficient at resolution level $|\lambda| = 3$, and (b) a wavelet coefficient at resolution level $|\lambda| = 5$.

In order to derive the minimax results, we define the minimax L^2-risk over the class of balls $\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$ as

$$\begin{aligned}
\mathcal{R}_\epsilon(\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)) &:= \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{\boldsymbol{f}\in\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} R(\hat{\boldsymbol{f}}_\epsilon,\boldsymbol{f}), \\
&= \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{\boldsymbol{f}\in\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} \mathbb{E}\|\hat{\boldsymbol{f}}_\epsilon - \boldsymbol{f}\|^2 \\
&= \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{\boldsymbol{f}\in\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} \left(\int_{T\times\mathcal{X}} \left| \hat{\boldsymbol{f}}_\epsilon(t,\underline{x}) - \boldsymbol{f}(t,\underline{x}) \right|^2 dt d\underline{x} \right),
\end{aligned}$$

where $\|\mathbf{g}\|$ is the L^2-norm of a function $\mathbf{g}$ defined on $T \times \mathcal{X}$ and the infimum is taken over all possible estimators $\hat{\boldsymbol{f}}_\epsilon$ (i.e., measurable functions) of $\boldsymbol{f}$, based on observations from model (1.1).

To present our results, for any $d \geq 1$ and any $s_1 > 0$ and $s_2 > 0$, we define $s > 0$ to be such that

$$\frac{1}{s} = \frac{1}{d+1}\left(\frac{d}{s_1} + \frac{1}{s_2}\right). \tag{4.4}$$

In what follows, we use the symbol C for a generic positive constant, independent of ϵ, which may take different values at different places. Moreover, in order to simplify the presentation of the results, and without loss of generality, we assume below that $j_0 = m_0 = 0$.

5 Minimax lower bound for the L^2-risk

The following statement provides the minimax lower bounds for the L^2-risk.

Theorem 1. *Let $A_1 > 0$ and $A_2 > 0$ be constants. Let $(A_\lambda)_{\lambda\in\Lambda}$ be a set of positive constants satisfying (4.3), and assume that there exists a positive constant $A > 0$ such that, for any $-1 \leq j < +\infty$ and $|\lambda| = j$,*

$$A_\lambda 2^{\frac{d}{2}(j+1)} \geq A. \tag{5.1}$$

Let $s_1 > 0$ and $s_2 > 0$ be the smoothness parameters in the space and time domains, respectively, such that $0 < s_1 < \tau_1$ and $0 < s_2 < \tau_2$, where τ_1 and τ_2 are the regularity parameters of the wavelet systems (ϕ, ψ) and $(\tilde{\phi}, \tilde{\psi})$, respectively. Assume that $1 \leq p_1, q_1 \leq +\infty$, $1 \leq p_2, q_2 \leq +\infty$ such that $s_1 + d(1/2 - 1/p_1) > 0$ and $s_2 + 1/2 - 1/p_2 > 0$, and let $s > 0$ satisfy (4.4). Then, there exists a constant $C > 0$ such that

$$\mathcal{R}_\epsilon(\mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)) \geq C\, \epsilon^{\frac{4s}{2s+d+1}},$$

for all sufficiently small $\epsilon > 0$.

Proof. The proof is based on the standard Assouad's cube technique (see, e.g., [Tsybakov, 2009], Chapter 2, Section 2.7.2). Consider the following test functions

$$\boldsymbol{f}_w(t, \underline{x}) = \mu_{j_1,m_2} \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} w_{\lambda,m,\ell} \tilde{\psi}_{m,\ell}(t)\psi_\lambda(\underline{x}), \quad t \in T, \quad \underline{x} \in \mathcal{X},$$

where $w = \left((w_{\lambda,m,\ell})_{|\lambda|=j, 0\leq \ell \leq 2^m-1}\right)_{j=-1,\ldots,j_1, m=-1,\ldots,m_2} \in \Omega := \{1, -1\}^{2^{j_1 d+m_2}}$, and μ_{j_1,m_2} is a positive sequence of reals satisfying the condition

$$\mu_{j_1,m_2} = c 2^{-\frac{1}{2}(m_2+1)} 2^{-\frac{d}{2}(j_1+1)} \min\left(2^{-(j_1+1)s_1}, 2^{-(m_2+1)s_2}\right), \tag{5.2}$$

for some constant $c > 0$ not depending on j_1 and m_2. Assume that c satisfies the condition

$$c \leq \min\left(A_1/(\|\tilde{\psi}_{-1}\|_\infty + K\|\tilde{\psi}\|_\infty), A\right), \tag{5.3}$$

where A is the constant satisfying inequality (5.1) and K is a constant that is proportional to the length support of $\tilde{\psi}$. Then, it easily follows that $\boldsymbol{f}_w \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)$ for any $w \in \Omega$. Indeed, for any $t \in T$,

$$\|\boldsymbol{f}_w(t,\cdot)\|^{s_1}_{p_1,q_1} = \left(\sum_{j=-1}^{j_1} 2^{j(s_1+d(1/2-1/p_1))q_1} \left(\sum_{|\lambda|=j} |\langle \boldsymbol{f}_w(t,\cdot), \psi_\lambda \rangle|^{p_1} \right)^{q_1/p_1} \right)^{1/q_1},$$

where

$$|\langle \boldsymbol{f}_w(t,\cdot), \psi_\lambda \rangle| \leq \mu_{j_1,m_2} \left(\|\tilde{\psi}_{-1}\|_\infty + \sum_{m=0}^{m_2} \sum_{\ell=0}^{2^m-1} |\tilde{\psi}_{m,\ell}(t)| \right).$$

Let us define the set

$$I_m(t) = \{0 \leq \ell \leq 2^m - 1 \ : \ \tilde{\psi}_{m,\ell}(t) \neq 0\}.$$

Since, the wavelet $\tilde{\psi}$ is compactly supported, one has that the cardinality of $I_m(t)$ is bounded by a constant $K > 0$ that is proportional to the length support of $\tilde{\psi}$. Thus, using the relation $\|\tilde{\psi}_\lambda\|_\infty = 2^{m/2}\|\tilde{\psi}\|_\infty$, we obtain that, for any $t \in T$,

$$\begin{aligned}
|\langle \boldsymbol{f}_w(t,\cdot), \psi_\lambda \rangle| &\leq \mu_{j_1,m_2} \left(\|\tilde{\psi}_{-1}\|_\infty + \sum_{m=0}^{m_2} \sum_{\ell \in I_m(t)} |\tilde{\psi}_{m,\ell}(t)| \right) \\
&\leq \mu_{j_1,m_2} \left(\|\tilde{\psi}_{-1}\|_\infty + \sum_{m=0}^{m_2} K 2^{m/2} \|\tilde{\psi}\|_\infty \right) \\
&\leq \mu_{j_1,m_2} \left(\|\tilde{\psi}_{-1}\|_\infty + K\|\tilde{\psi}\|_\infty 2^{\frac{1}{2}(m_2+1)} \right) \\
&\leq \mu_{j_1,m_2} \left(\|\tilde{\psi}_{-1}\|_\infty + K\|\tilde{\psi}\|_\infty \right) 2^{\frac{1}{2}(m_2+1)}.
\end{aligned}$$

Therefore, by the definition of μ_{j_1,m_2} given in (5.2), it follows that

$$\begin{aligned}
\sup_{t\in T}\{\|\boldsymbol{f}_w(t,\cdot)\|^{s_1}_{p_1,q_1}\} &\leq (\|\tilde{\psi}_{-1}\|_\infty + K\|\tilde{\psi}\|_\infty)\mu_{j_1,m_2} 2^{\frac{1}{2}(m_2+1)} 2^{(j_1+1)(s_1+d/2)} \\
&\leq c(\|\tilde{\psi}_{-1}\|_\infty + K\|\tilde{\psi}\|_\infty). \qquad (5.4)
\end{aligned}$$

Now, define, for each $\lambda \in \Lambda$ and $t \in T$,

$$\boldsymbol{\alpha}_{w,\lambda}(t) = \mu_{j_1,m_2} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} w_{\lambda,m,\ell} \tilde{\psi}_{m,\ell}(t),$$

with $w = \big((w_{\lambda,m,\ell})_{|\lambda|=j, 0\leq \ell \leq 2^m-1} \big)_{j=-1,\ldots,j_1, m=-1,\ldots,m_2}$ and μ_{j_1,m_2} are as given above. On noting that

$$|\langle \boldsymbol{\alpha}_{w,\lambda}, \tilde{\psi}_{m,l} \rangle| \leq \mu_{j_1,m_2} |w_{\lambda,m,l}|,$$

and using the definition of μ_{j_1,m_2} given in (5.2) and the inequality (5.1), we obtain that, for any $|\lambda| = j$ with $-1 \leq j \leq j_1$,

$$\begin{aligned}
\|\boldsymbol{\alpha}_{w,\lambda}\|^{s_2}_{p_2,q_2} &\leq \mu_{j_1,m_2} 2^{(m_2+1)(s_2+1/2)} \\
&\leq c 2^{-\frac{d}{2}(j_1+1)} \leq cA_\lambda/A. \qquad (5.5)
\end{aligned}$$

Hence, using the inequalities (5.4) and (5.5), it follows that the condition (5.3) is sufficient to imply that $\boldsymbol{f}_w \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)$.

In the rest of the proof, we will thus assume that condition (5.3) holds. Furthermore, we use the notation $\mathbb{E}_{\boldsymbol{f}_w}$ to denote expectation with respect to the distribution $\mathbb{P}_{\boldsymbol{f}_w}$ of the random process Y in model (1.1) under the hypothesis that $\boldsymbol{f} = \boldsymbol{f}_w$.

The minimax risk $\mathcal{R}_\epsilon(\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2))$ can be bounded from below as follows

$$\begin{aligned}\mathcal{R}_\epsilon := \mathcal{R}_\epsilon(\mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)) &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{w\in\Omega} R(\hat{\boldsymbol{f}}_\epsilon, \boldsymbol{f}_w) \\ &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{w\in\Omega} \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} \mathbb{E}_{\boldsymbol{f}_w} \left| \hat{\alpha}^\epsilon_{\lambda,m,\ell} - \mu_{j_1,m_2} w_{\lambda,m,\ell} \right|^2\end{aligned}$$

where

$$\hat{\alpha}^\epsilon_{\lambda,m,\ell} = \int_{T\times\mathcal{X}} \hat{\boldsymbol{f}}_\epsilon(t,\underline{x}) \tilde{\psi}_{m,\ell}(t) \psi_\lambda(\underline{x}) dt d\underline{x}.$$

Then, define

$$\hat{w}^\epsilon_{\lambda,m,\ell} := \arg\min_{v\in\{-1,1\}} \left| \hat{\alpha}^\epsilon_{\lambda,m,\ell} - \mu_{j_1,m_2} v \right|,$$

and remark that the triangular inequality and the definition of $\hat{w}^\epsilon_{\lambda,m,\ell}$ imply that

$$\mu_{j_1,m_2} \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w_{\lambda,m,\ell} \right| \leq 2 \left| \hat{\alpha}^\epsilon_{\lambda,m,\ell} - \mu_{j_1,m_2} w_{\lambda,m,\ell} \right|,$$

which yields

$$\begin{aligned}\mathcal{R}_\epsilon &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \sup_{w\in\Omega} \frac{\mu^2_{j_1,m_2}}{4} \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} \mathbb{E}_{\boldsymbol{f}_w} \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w_{\lambda,m,\ell} \right|^2 \\ &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \frac{\mu^2_{j_1,m_2}}{4} \frac{1}{\#\Omega} \sum_{w\in\Omega} \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} \mathbb{E}_{\boldsymbol{f}_w} \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w_{\lambda,m,\ell} \right|^2.\end{aligned}$$

Replacing the sums $\sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1}$ by $\sum_{\lambda,m,\ell}$ (to simplify the notation), for any λ, m, ℓ and $w \in \Omega$, define the vector $w^{(\lambda,m,\ell)} \in \Omega$ having all its components equal to w expect the (λ,m,ℓ)-th element. Let $\#A$ denote the cardinality of a finite set A. Then

$$\begin{aligned}\mathcal{R}_\epsilon &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \frac{\mu^2_{j_1,m_2}}{4} \frac{1}{\#\Omega} \sum_{\lambda,m,\ell} \sum_{w\in\Omega\,:\,w_{\lambda,m,\ell}=1} \left(\mathbb{E}_{\boldsymbol{f}_w} \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w_{\lambda,m,\ell} \right|^2 + \mathbb{E}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}} \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w^{(\lambda,m,\ell)}_{\lambda,m,\ell} \right|^2 \right) \\ &\geq \inf_{\hat{\boldsymbol{f}}_\epsilon} \frac{\mu^2_{j_1,m_2}}{4} \frac{1}{\#\Omega} \sum_{\lambda,m,\ell} \sum_{w\in\Omega\,:\,w_{\lambda,m,\ell}=1} \mathbb{E}_{\boldsymbol{f}_w} \left(\left| \hat{w}^\epsilon_{\lambda,m,\ell} - w_{\lambda,m,\ell} \right|^2 + \left| \hat{w}^\epsilon_{\lambda,m,\ell} - w^{(\lambda,m,\ell)}_{\lambda,m,\ell} \right|^2 \frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y) \right).\end{aligned}$$

Since $w^{(\lambda,m,\ell)}_{\lambda,m,\ell} = -w_{\lambda,m,\ell}$ and $\hat{w}^\epsilon_{\lambda,m,\ell} \in \{-1,1\}$, one finally obtains that, for any $0 < \delta < 1$,

$$\begin{aligned}\mathcal{R}_\epsilon &\geq \mu^2_{j_1,m_2} \frac{1}{\#\Omega} \sum_{\lambda,m,\ell} \sum_{w\in\Omega\,:\,w_{\lambda,m,\ell}=1}^{2} \mathbb{E}_{\boldsymbol{f}_w} \left(\min\left(1, \frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y) \right) \right) \\ &\geq \delta \mu^2_{j_1,m_2} \frac{1}{\#\Omega} \sum_{\lambda,m,\ell} \sum_{w\in\Omega\,:\,w_{\lambda,m,\ell}=1}^{2} \mathbb{P}_{\boldsymbol{f}_w} \left(\frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y) > \delta \right). \qquad (5.6)\end{aligned}$$

Thanks to the multiparameter Girsanov's formula, one has that, under the hypothesis that $\boldsymbol{f} = \boldsymbol{f}_w$ in model (1.1),

$$\log\left(\frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y)\right) = \epsilon^{-1}\int_{T\times\mathcal{X}}(\boldsymbol{f}_{w^{(\lambda,m,\ell)}} - \boldsymbol{f}_w)(t,\underline{x})dW(t,\underline{x}) - \frac{\epsilon^{-2}}{2}\int_{T\times\mathcal{X}}(\boldsymbol{f}_{w^{(\lambda,m,\ell)}} - \boldsymbol{f}_w)^2(t,\underline{x})dtd\underline{x}$$

Therefore, the random variable

$$Z_{\lambda,m,\ell} := \log\left(\frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y)\right)$$

is Gaussian with mean $\theta = -\frac{\epsilon^{-2}}{2}\mu^2_{j_1,m_2}$ and variance $\sigma^2 = \epsilon^{-2}\mu^2_{j_1,m_2}$, that do not depend on (λ, m, ℓ).

Now, let $s > 0$ satisfy (4.4). Define $j_1 = j_1(\epsilon)$ and $m_2 = m_2(\epsilon)$ as

$$2^{(j_1(\epsilon)+1)} = \lfloor \epsilon^{-\frac{2s}{(2s+d+1)s_1}} \rfloor \quad \text{and} \quad 2^{(m_2(\epsilon)+1)} = \lfloor \epsilon^{-\frac{2s}{(2s+d+1)s_2}} \rfloor. \tag{5.7}$$

Thanks to (5.2), it follows that there exists $c_1 > 0$ such that

$$\epsilon^{-2}\mu^2_{j_1(\epsilon),m_2(\epsilon)} \leq c_1$$

for all sufficiently small $\epsilon > 0$. Hence, $Z_{\lambda,m,\ell} \sim N(\theta, \sigma^2)$ with $|\theta| \leq c_1/2$ and $\sigma^2 \leq c_1$ which implies that there exist $\gamma > 0$ and $0 < \delta < 1$, that do not depend on (λ, m, ℓ) and ϵ, such that for all sufficiently small $\epsilon > 0$

$$\mathbb{P}_{\boldsymbol{f}_w}\left(\frac{d\mathbb{P}_{\boldsymbol{f}_{w^{(\lambda,m,\ell)}}}}{d\mathbb{P}_{\boldsymbol{f}_w}}(Y) > \log(\delta)\right) \geq \gamma.$$

Hence, inserting the above inequality into (5.6), it implies that

$$\mathcal{R}_\epsilon \geq \frac{1}{2}\delta\gamma\mu^2_{j_1(\epsilon),m_2(\epsilon)}\, 2^{(j_1(\epsilon)+1)d+(m_2(\epsilon)+1)}.$$

Using the expressions of $j_1(\epsilon)$ and $m_2(\epsilon)$ given in (5.7), together with (4.4) and (5.2), we finally obtain that there exists a constant $C > 0$, that does not depend on ϵ, such that

$$\mathcal{R}_\epsilon \geq C\epsilon^{\frac{4s}{2s+d+1}},$$

for all sufficiently small $\epsilon > 0$, thus completing the proof of the theorem.

□

6 Minimax upper bound for the L^2-risk

We now provide minimax upper bounds for the L^2-risk. This will accomplished by constructing appropriate estimators of $\boldsymbol{f}(t, \underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$, in the sequence space model.

6.1 The sequence space model

The suggested estimators in the following sections, will be constructed on the sequence space. Let us first recall that (see, e.g., [Chow et al., 2001]) (1.1) must be understood in the following sense: for any $\mathbf{g} \in \mathbb{L}^2(T \times \mathcal{X})$,

$$\int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})dY(t,\underline{x}) = \int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})\boldsymbol{f}(t,\underline{x})dtd\underline{x} + \epsilon \int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})dW(t,\underline{x})$$

so that the integrand of "the data" $dY(t,\underline{x})$ with respect to $\mathbf{g}(t,\underline{x})$ is a random variable that is normally distributed with mean

$$\mathbb{E}\left(\int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})dY(t,\underline{x})\right) = \int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})\boldsymbol{f}(t,\underline{x})dtd\underline{x}$$

and variance

$$\mathrm{Var}\left(\int_{T\times\mathcal{X}} \mathbf{g}(t,\underline{x})dY(t,\underline{x})\right) = \epsilon^2 \int_{T\times\mathcal{X}} |\mathbf{g}(t,\underline{x})|^2 dtd\underline{x}.$$

Moreover, for any $g_1, g_2 \in \mathbb{L}^2(T \times \mathcal{X})$

$$\mathbb{E}\left(\int_{T\times\mathcal{X}} \mathbf{g}_1(t,\underline{x})dW(t,\underline{x}) \int_{T\times\mathcal{X}} \mathbf{g}_2(t,\underline{x})dW(t,\underline{x})\right) = \int_{T\times\mathcal{X}} g_1(t,\underline{x})g_2(t,\underline{x})dtd\underline{x}.$$

Hence, in view of the above and using the tensor product wavelet basis constructed in Section 3, noisy observations of the coefficients $\tilde{\alpha}_{\lambda,m,\ell}$ are thus obtained through the following sequence model

$$\begin{aligned} y_{\lambda,m,\ell} &= \int_{T\times\mathcal{X}} \tilde{\psi}_{m,\ell}(t)\psi_\lambda(\underline{x})dY(t,\underline{x}) \\ &= \tilde{\alpha}_{\lambda,m,\ell} + \epsilon\, z_{\lambda,m,\ell},\ \lambda \in \Lambda,\ m \geq -1,\ \ell = 0,1,\ldots,2^m-1, \end{aligned} \tag{6.1}$$

where the $z_{\lambda,m,\ell}$'s are independent and identically distributed (i.i.d.) standard Gaussian random variables, i.e., Gaussian random variables with zero mean and variance 1.

6.2 Linear and non-adaptive estimator

Consider the sequence space model (6.2). Let $j_1 > 0$ and $m_2 > 0$ be integers (smoothing parameters). We consider the following non-adaptive wavelet projection (linear) estimator of $\boldsymbol{f}(t,\underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$, that is

$$\hat{\boldsymbol{f}}^l_{j_1,m_2}(t,\underline{x}) = \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} y_{\lambda,m,\ell}\tilde{\psi}_{m,\ell}(t)\psi_\lambda(\underline{x}), \quad t \in T, \quad \underline{x} \in \mathcal{X}. \tag{6.2}$$

Define the L^2-risk of $\hat{\boldsymbol{f}}^l_{j_1,m_2}$ as

$$\begin{aligned} R(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) &= \mathbb{E}\|\hat{\boldsymbol{f}}^l_{j_1,m_2} - \boldsymbol{f}\|^2_{\mathbb{L}^2(T\times\mathcal{X})} \\ &= \mathbb{E}\Big(\int_{T\times\mathcal{X}} \left|\hat{\boldsymbol{f}}^l_{j_1,m_2}(t,\underline{x}) - \boldsymbol{f}(t,\underline{x})\right|^2 dtd\underline{x}\Big). \end{aligned}$$

The following statement provides the minimax upper bounds for the L^2-risk of the non-adaptive (linear) wavelet estimator $\hat{\boldsymbol{f}}^l_{j_1,m_2}$ given in (6.2).

Theorem 2. *Let $A_1 > 0$ and $A_2 > 0$ be constants. Let $s_1 > 0$ and $s_2 > 0$ be the smoothness parameters in the space and time domains, respectively, such that $0 < s_1 < \tau_1$ and $0 < s_2 < \tau_2$, where τ_1 and τ_2 are the regularity parameters of the wavelet systems (ϕ, ψ) and $(\tilde{\phi}, \tilde{\psi})$, respectively. Assume that $2 \leq p_1, q_1 \leq +\infty$, $2 \leq p_2, q_2 \leq +\infty$ such that $s_1 + d(1/2 - 1/p_1) > 0$ and $s_2 + 1/2 - 1/p_2 > 0$, and let $s > 0$ satisfy (4.4). Consider the linear estimator $\hat{\boldsymbol{f}}^l_{j_1,m_2}$ given in (6.2), and define $j_1 = j_1(\epsilon)$ and $m_2 = m_2(\epsilon)$ such that*

$$2^{(j_1(\epsilon)+1)} = \lfloor \epsilon^{-\frac{2s}{(2s+d+1)s_1}} \rfloor \quad \textit{and} \quad 2^{(m_2(\epsilon)+1)} = \lfloor \epsilon^{-\frac{2s}{(2s+d+1)s_2}} \rfloor. \tag{6.3}$$

Then, there exists a constant $C > 0$ such that

$$\sup_{\boldsymbol{f} \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} R(\hat{\boldsymbol{f}}_{j_1(\epsilon),m_2(\epsilon)}, \boldsymbol{f}) \leq C\, \epsilon^{\frac{4s}{2s+d+1}},$$

for all sufficiently small $\epsilon > 0$.

Proof. Let us write the usual bias-variance decomposition of the L^2-risk as

$$R(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) = B(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) + V(\hat{\boldsymbol{f}}^l_{j_1,m_2}),$$

with

$$B(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) = \|\mathbb{E}\hat{\boldsymbol{f}}^l_{j_1,m_2} - \boldsymbol{f}\|^2 \quad \text{and} \quad V(\hat{\boldsymbol{f}}^l_{j_1,m_2}) = \mathbb{E}\|\hat{\boldsymbol{f}}^l_{j_1,m_2} - \mathbb{E}\hat{\boldsymbol{f}}^l_{j_1,m_2}\|^2.$$

Obviously,

$$\begin{aligned} V(\hat{\boldsymbol{f}}^l_{j_1,m_2}) &= \epsilon^2 \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{\ell=0}^{2^m-1} \mathbb{E}|z_{\lambda,m,\ell}|^2 \\ &= \epsilon^2 2^{(j_1+1)d+m_2+1}, \end{aligned} \tag{6.4}$$

and

$$B(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) = B_1(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) + B_2(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}),$$

where

$$\begin{aligned} B_1(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) &= \sum_{j=j_1+1}^{+\infty} \sum_{|\lambda|=j} \sum_{m=-1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \\ &= \sum_{j=j_1+1}^{\infty} \sum_{|\lambda|=j} \int_T |\boldsymbol{\alpha}_\lambda(t)|^2\, dt \\ &= \int_T \sum_{j=j_1+1}^{\infty} \sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2\, dt \end{aligned}$$

and

$$\begin{aligned} B_2(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) &= \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=m_2+1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \\ &\leq \sum_{j=-1}^{+\infty} \sum_{|\lambda|=j} \sum_{m=m_2+1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2. \end{aligned}$$

By Lemma 1, there exists a constant $K_2 > 0$ (only depending on s_2 and p_2) such that

$$\sum_{m=m_2+1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \leq K_2 A_\lambda^2 2^{-2(m_2+1)s_2}.$$

Thus, using (4.3), it follows that

$$B_2(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) \leq K_2 A_2^2 2^{-2(m_2+1)s_2}. \tag{6.5}$$

Moreover, by Lemma 1, there exists a constant $K_1 > 0$ (only depending on s_1 and p_1) such that

$$\sum_{j=j_1+1}^{+\infty} \sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2 \leq K_1 A_1^2 2^{-2(j_1+1)s_1},$$

This implies that

$$B(\hat{\boldsymbol{f}}^l_{j_1,m_2}, \boldsymbol{f}) \leq K_1 A_1^2 2^{-2(j_1+1)s_1} + K_2 A_2^2 2^{-2(m_2+1)s_2}. \tag{6.6}$$

Therefore, by combining (6.4), (6.5) and (6.6), we arrive at

$$R(\hat{\boldsymbol{f}}_{j_1(\epsilon),m_2(\epsilon)}, \boldsymbol{f}) \leq K_1 A_1^2 2^{-2(j_1(\epsilon)+1)s_1} + K_2 A_2^2 2^{-2(m_2(\epsilon)+1)s_2} + \epsilon^2 2^{(j_1(\epsilon)+1)d+(m_2(\epsilon)+1)}.$$

By taking into account the expressions of $j_1(\epsilon)$ and $m_2(\epsilon)$ given in (6.3), together with (4.4), we finally obtain that there exists a constant $C > 0$, that does not depend on ϵ, such that

$$\sup_{\boldsymbol{f} \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} R(\hat{\boldsymbol{f}}_{j_1(\epsilon),m_2(\epsilon)}, \boldsymbol{f}) \leq C\, \epsilon^{\frac{4s}{2s+d+1}},$$

for all sufficiently small $\epsilon > 0$, thus completing the proof of the theorem. □

The choice of the resolution levels j_1 and m_2 depends on the unknown smoothness parameters s_1 and s_2 in the space and time domains, respectively. The linear estimator $\hat{\boldsymbol{f}}^l_{j_1,m_2}$ defined in (6.2) is thus called non-adaptive (with respect to s_1 and s_2) and is of limited interest in practical applications. Moreover, the results of Theorem 2 are only suited to model d-dimensional functions $\boldsymbol{f}(t,\cdot)$ belonging to the space $B^{s_1}_{p_1,q_1}(A_1)$ with $2 \leq p_1, q_1 \leq +\infty$, uniformly over $t \in T$. However, such Besov spaces are not suited to model spatially inhomogeneous multivariate functions.

In the following section, we thus consider the problem of constructing an adaptive non-linear estimator that is optimal (in the minimax sense) over Besov balls $\mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)$ with $1 \leq p_1, q_1 \leq +\infty$ and $1 \leq p_2, q_2 \leq +\infty$.

6.3 Non-linear and adaptive estimation

Consider the sequence space model (6.2). For each $\lambda \in \Lambda$, we divide the wavelet coefficients $\alpha_{\lambda,m,\ell}$ at each resolution level $-1 \leq m < +\infty$ into blocks of length $L_\epsilon = 1 + \lfloor \log(\epsilon^{-2}) \rfloor$. Let A_m and U_{mr} be the following sets of indices

$$\begin{aligned} A_m &= \left\{ r \mid r = 1, 2, \ldots, \frac{2^m}{L_\epsilon} \right\}, \\ U_{mr} &= \left\{ \ell \mid \ell = 0, 1, \ldots, 2^m - 1; (r-1)L_\epsilon \leq \ell \leq rL_\epsilon - 1 \right\}. \end{aligned}$$

Now, we define

$$B_{\lambda,m,r} = \sum_{\ell \in U_{mr}} \alpha^2_{\lambda,m,\ell} \quad \text{and} \quad \hat{B}_{\lambda,m,r} = \sum_{\ell \in U_{mr}} y^2_{\lambda,m,\ell}. \tag{6.7}$$

We consider an adaptive wavelet block-thresholding (non-linear) estimator $\boldsymbol{f}(t,\underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$, that is

$$\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}(t,\underline{x}) = \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{r \in A_m} \sum_{\ell \in U_{mr}} y_{\lambda,m,\ell} \mathbb{1}_{\left\{\hat{B}_{\lambda,m,r} \geq t_{\epsilon,\delta}\right\}} \tilde{\psi}_{m,\ell}(t)\psi_\lambda(\underline{x}), \quad t \in T, \quad \underline{x} \in \mathcal{X}, \tag{6.8}$$

where $\mathbb{1}_A$ is the indicator function of the set A, and the resolution levels j_1 and m_2, and the threshold $t_{\epsilon,\delta}$, will be defined below.

Define the L^2-risk of $\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}$ as

$$\begin{aligned} R(\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}, \boldsymbol{f}) &= \mathbb{E}\|\hat{\boldsymbol{f}}^{nl}_{j_1,m_2} - \boldsymbol{f}\|^2_{\mathbb{L}^2(T\times\mathcal{X})} \\ &= \mathbb{E}\bigg(\int_{T\times\mathcal{X}} \Big|\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}(t,\underline{x}) - \boldsymbol{f}(t,\underline{x})\Big|^2 dt d\underline{x}\bigg). \end{aligned}$$

The following statement provides the minimax upper bounds for the L^2-risk of the adaptive (non-linear) wavelet estimator $\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}$ given in (6.8).

Theorem 3. *Let $A_1 > 0$ and $A_2 > 0$ be constants. Let $s_1 > 0$ and $s_2 > 0$ be the smoothness parameters in the space and time domains, respectively, such that $0 < s_1 < \tau_1$ and $0 < s_2 < \tau_2$, where τ_1 and τ_2 are the regularity parameters of the wavelet systems (ϕ,ψ) and $(\tilde{\phi},\tilde{\psi})$, respectively. Assume that $1 \leq p_1, q_1 \leq +\infty$, $1 \leq p_2, q_2 \leq +\infty$ such that $s_1 + d(1/2 - 1/p_1) > 0$ and $s_2 + 1/2 - 1/p_2 > 0$ if $2 \leq p_1, q_1 \leq +\infty$ and $2 \leq p_2, q_2 \leq +\infty$, respectively, and $s_1 \geq d/p_1$ and $s_2 \geq 1/p_2$ if $1 \leq p_1, q_1 < 2$ and $1 \leq p_2, q_2 < 2$, respectively. Let also $s > 0$ satisfy (4.4). Consider the non-linear estimator $\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}$ given in (6.8), and define $j_1 = j_1(\epsilon)$ and $m_2 = m_2(\epsilon)$ as*

$$2^{(j_1(\epsilon)+1)} = \lfloor \epsilon^{-2} \rfloor \quad \text{and} \quad 2^{(m_2(\epsilon)+1)} = \lfloor \epsilon^{-2} \rfloor. \tag{6.9}$$

Define the threshold

$$t_{\epsilon,\delta} = \delta^2 \epsilon^2 L_\epsilon,$$

for some $\delta > 2(2\sqrt{2}+1)$. Then, there exists a constant $C > 0$ such that

$$\sup_{\boldsymbol{f} \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1,A_2)} R(\hat{\boldsymbol{f}}^{nl}_{j_1(\epsilon),m_2(\epsilon)}, \boldsymbol{f}) \leq C\, \epsilon^{\frac{4s}{2s+d+1}},$$

for all sufficiently small $\epsilon > 0$.

Proof. From Parseval's equality, we can decompose the L^2-risk of $\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}$ as follows

$$R(\hat{\boldsymbol{f}}^{nl}_{j_1,m_2}, \boldsymbol{f}) = B_1 + B_2 + R_1 + R_2,$$

where

$$
\begin{aligned}
B_1 &= \sum_{j=j_1+1}^{+\infty} \sum_{|\lambda|=j} \sum_{m=-1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 = \int_T \sum_{j=j_1+1}^{\infty} \sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2 \, dt, \\
B_2 &= \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=m_2+1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2, \\
R_1 &= \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{r\in A_m} \sum_{\ell\in U_{mr}} \mathbb{E}\left(|y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \, \mathbb{1}_{\left\{\hat{B}_{\lambda,m,r} \geq t_{\epsilon,\delta}\right\}} \right), \\
R_2 &= \sum_{j=-1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \sum_{r\in A_m} \sum_{\ell\in U_{mr}} \mathbb{E}\left(|\tilde{\alpha}_{\lambda,m,\ell}|^2 \, \mathbb{1}_{\left\{\hat{B}_{\lambda,m,r} < t_{\epsilon,\delta}\right\}} \right).
\end{aligned}
$$

To bound the risk, we need to control the terms B_1, B_2, R_1 and R_2. Let $p_1' = \min(p_1, 2)$ and $p_2' = \min(p_2, 2)$. Define also $s_1' = s_1 + d(1/2 - 1/p_1')$ and $s_2' = s_2 + 1/2 - 1/p_2'$.

By Lemma 1, there exists a constant $K_1' > 0$, only depending on s_1 and p_1, such that

$$
\sum_{j=j_1+1}^{+\infty} \sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2 \leq K_1' A_1^2 2^{-2(j_1+1)s_1'}
$$

implying that

$$
B_1 \leq K_1' A_1^2 2^{-2(j_1+1)s_1'}.
$$

Also, by Lemma 1, there exists a constant $K_2' > 0$, only depending on s_2 and p_2, such that

$$
\sum_{m=m_2+1}^{+\infty} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \leq K_2' A_\lambda^2 2^{-2(m_2+1)s_2'},
$$

implying, in view of equation (4.3), that

$$
B_2 \leq K_2' A_2^2 2^{-2(m_2+1)s_2'}.
$$

Consider the case $2 \leq p_1 \leq +\infty$ implying that $s_1' = s_1$. Thanks to the definitions of $j_1(\epsilon)$ given in (6.9) and s given in (4.4), we obtain that

$$
B_1 = o\left(\epsilon^{\frac{4s}{2s+d+1}}\right).
$$

In the case $1 \leq p_1 < 2$, the condition $s_1 \geq d/p_1$, the definitions of $j_1(\epsilon)$ given in (6.9) and s given (4.4) also imply that

$$
B_1 = o\left(\epsilon^{\frac{4s}{2s+d+1}}\right).
$$

Consider the case $2 \leq p_2 \leq +\infty$ implying that $s_2' = s_2$. Thanks to the definitions of $m_2(\epsilon)$ given in (6.9) and s given in (4.4), we obtain that

$$
B_2 = o\left(\epsilon^{\frac{4s}{2s+d+1}}\right).
$$

In the case $1 \leq p_2 < 2$, the condition $s_2 \geq 1/p_2$, the definitions of $m_2(\epsilon)$ given in (6.9) and s given in (4.4) also imply that

$$B_2 = o\Big(\epsilon^{\frac{4s}{2s+d+1}}\Big)\,.$$

Let us now write R_1 and R_2 as the sum of two terms

$$R_1 = R_{1,1} + R_{1,2}$$

where

$$\begin{aligned}
R_{1,1} &= \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}\sum_{r\in A_m}\sum_{\ell\in U_{mr}} \mathbb{E}\left(|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2\,\mathbb{1}_{\left\{\sum_{\ell\in U_{mr}}|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2\geq\frac{1}{4}t_{\epsilon,\delta}\right\}}\right),\\
R_{1,2} &= \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}\sum_{r\in A_m}\sum_{\ell\in U_{mr}} \mathbb{E}\left(|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2\right)\mathbb{1}_{\left\{B_{\lambda,m,r}>\frac{1}{4}t_{\epsilon,\delta}\right\}},\\
R_{2,1} &= \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}\sum_{r\in A_m}\sum_{\ell\in U_{mr}} \mathbb{E}\left(|\tilde{\alpha}_{\lambda,m,\ell}|^2\,\mathbb{1}_{\left\{\sum_{\ell\in U_{mr}}|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2\geq\frac{1}{4}t_{\epsilon,\delta}\right\}}\right),\\
R_{2,2} &= \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}\sum_{r\in A_m}\sum_{\ell\in U_{mr}} |\tilde{\alpha}_{\lambda,m,\ell}|^2\,\mathbb{1}_{\left\{B_{\lambda,m,r}<\frac{5}{2}t_{\epsilon,\delta}\right\}},
\end{aligned}$$

where we have used the inequality $|y_{\lambda,m,\ell}|^2 \leq 2\,|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2 + 2\,|\tilde{\alpha}_{\lambda,m,\ell}|^2$.

Let us first give an upper bound for $\Delta_1 = R_{1,1} + R_{2,1}$ as follows. Using Cauchy-Schwarz's inequality, moments properties of Gaussian random variables, Lemma 1 and Lemma 2, we have

$$\begin{aligned}
\Delta_1 &\leq \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}\sum_{r\in A_m}\sum_{\ell\in U_{mr}}\left(\mathbb{E}\left(|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^4\right)^{1/2} + |\tilde{\alpha}_{\lambda,m,\ell}|^2\right)\\
&\quad\times\left(\mathbb{P}\left(\sum_{\ell\in U_{mr}}|y_{\lambda,m,\ell}-\tilde{\alpha}_{\lambda,m,\ell}|^2\geq\frac{1}{4}t_{\epsilon,\delta}\right)\right)^{1/2}\\
&\leq \left(\sqrt{3}\,2^{(j_1+1)d+(m_2+1)}\epsilon^2 + \sum_{j=-1}^{j_1}\sum_{|\lambda|=j}\sum_{m=-1}^{m_2}K_2'A_\lambda^2 2^{-ms_2'}\right)\epsilon^{\frac{1}{2}(\delta/2-1)^2}\\
&= O\Big(2^{(j_1+1)d+(m_2+1)}\epsilon^{2+\frac{1}{2}(\delta/2-1)^2} + \epsilon^{\frac{1}{2}(\delta/2-1)^2}\Big)\\
&= O\big(\epsilon^{-2}\big)\,,
\end{aligned}$$

where we have used the assumption that $\delta > 2(2\sqrt{2}+1)$.

Now, let $\Delta_2 = R_{1,2} + R_{2,2}$. Let j_0 and m_0 be defined as

$$2^{j_0+1} = \lfloor\epsilon^{-\frac{2s}{(2s+d+1)s_1'}}\rfloor \quad\text{ and }\quad 2^{m_0+1} = \lfloor\epsilon^{-\frac{2s}{(2s+d+1)s_2'}}\rfloor,$$

where $s_1' = s_1 + d(1/2 - 1/p_1')$ and $s_2' = s_2 + 1/2 - 1/p_2'$. Note that $-1 \leq j_0 < j_1$ and $-1 \leq m_0 < m_2$ for all sufficiently small $\epsilon > 0$. Then, Δ_2 can be partitioned as $\Delta_2 = \Delta_{2,1} + \Delta_{2,2}$

where the first component $\Delta_{2,1}$ is calculated over the indices $-1 \leq j \leq j_0$ and $-1 \leq m \leq m_0$, namely

$$\Delta_{2,1} = \sum_{j=-1}^{j_0} \sum_{|\lambda|=j} \sum_{m=-1}^{m_0} \left[\sum_{\ell=0}^{2^m-1} \mathbb{E}\left(|y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \right) \mathbb{1}_{\left\{ B_{\lambda,m,r} > \frac{1}{4} t_{\epsilon,\delta} \right\}} + \sum_{r \in A_m} B_{\lambda,m,r} \mathbb{1}_{\left\{ B_{\lambda,m,r} < \frac{5}{2} t_{\epsilon,\delta} \right\}} \right],$$

and the second component $\Delta_{2,2}$ is calculated over the remaining indices, namely

$$\begin{aligned}
\Delta_{2,2} \;=\; & \sum_{j=j_0+1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \left[\sum_{\ell=0}^{2^m-1} \mathbb{E}\left(|y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \right) \mathbb{1}_{\left\{ B_{\lambda,m,r} > \frac{1}{4} t_{\epsilon,\delta} \right\}} + \sum_{r \in A_m} B_{\lambda,m,r} \mathbb{1}_{\left\{ B_{\lambda,m,r} < \frac{5}{2} t_{\epsilon,\delta} \right\}} \right] \\
& + \sum_{j=-1}^{j_0} \sum_{|\lambda|=j} \sum_{m=m_0+1}^{m_2} \left[\sum_{\ell=0}^{2^m-1} \mathbb{E}\left(|y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \right) \mathbb{1}_{\left\{ B_{\lambda,m,r} > \frac{1}{4} t_{\epsilon,\delta} \right\}} + \sum_{r \in A_m} B_{\lambda,m,r} \mathbb{1}_{\left\{ B_{\lambda,m,r} < \frac{5}{2} t_{\epsilon,\delta} \right\}} \right].
\end{aligned}$$

Let us first give an upper bound for $\Delta_{2,1}$ as follows

$$\begin{aligned}
\Delta_{2,1} \;&\leq\; \sum_{j=-1}^{j_0} \sum_{|\lambda|=j} \sum_{m=-1}^{m_0} \left[\sum_{\ell=0}^{2^m-1} \mathbb{E}\left(|y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \right) + \sum_{r \in A_m} \frac{5}{2} t_{\epsilon,\delta} \right] \\
&=\; O\Big(2^{(j_0+1)d + (m_0+1)} \epsilon^2 \Big) \\
&=\; O\Big(\epsilon^{\frac{4s}{2s+d+1}} \Big),
\end{aligned}$$

where we have used the moments properties of Gaussian random variables, and the fact that the blocks A_m are of length L_ϵ.

Now, we compute an upper bound for $\Delta_{2,2}$.We have

$$\begin{aligned}
\Delta_{2,2} \;\leq\; & \sum_{j=j_0+1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \left[\sum_{r \in A_m} \sum_{\ell \in U_{mr}} \epsilon^2 \mathbb{1}_{\left\{ B_{\lambda,m,r} > \frac{1}{4} t_{\epsilon,\delta} \right\}} + \sum_{r \in A_m} B_{\lambda,m,r} \right] \\
& + \sum_{j=-1}^{j_0} \sum_{|\lambda|=j} \sum_{m=m_0+1}^{m_2} \left[\sum_{r \in A_m} \sum_{\ell \in U_{mr}} \epsilon^2 \mathbb{1}_{\left\{ B_{\lambda,m,r} > \frac{1}{4} t_{\epsilon,\delta} \right\}} + \sum_{r \in A_m} B_{\lambda,m,r} \right].
\end{aligned}$$

Noticing that $\delta \geq 2$, we see that $\sum_{r \in A_m} \sum_{\ell \in U_{mr}} \epsilon^2 \leq \frac{1}{4} t_{\epsilon,\delta}$, which implies

$$\begin{aligned}
\Delta_{2,2} \;\leq\; & 2 \sum_{j=j_0+1}^{j_1} \sum_{|\lambda|=j} \sum_{m=-1}^{m_2} \left[\sum_{r \in A_m} B_{\lambda,m,r} \right] \\
& + 2 \sum_{j=-1}^{j_0} \sum_{|\lambda|=j} \sum_{m=m_0+1}^{m_2} \left[\sum_{r \in A_m} B_{\lambda,m,r} \right].
\end{aligned}$$

Then, noticing that $\sum_{r \in A_m} B_{\lambda,m,r} = \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2$, by Lemma 1 we have

$$\begin{aligned}
\Delta_{2,2} &= O\left(\sum_{j=j_0+1}^{j_1} \sum_{|\lambda|=j} \int_T |\boldsymbol{\alpha}_\lambda(t)|^2 \, dt\right) + O\left(\sum_{j=-1}^{j_0} \sum_{|\lambda|=j} A_\lambda^2 \sum_{m=m_0+1}^{m_2} \sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2\right) \\
&= O\left(\sum_{j=j_0+1}^{j_1} 2^{-2js'_1}\right) + O\left(\sum_{j=-1}^{j_0} \sum_{|\lambda|=j} A_\lambda^2 \sum_{m=m_0+1}^{m_2} 2^{-2ms'_2}\right) \\
&= O\left(2^{-2(j_0+1)s'_1}\right) + O\left(2^{-2(m_0+1)s'_2}\right) = O\left(\epsilon^{\frac{4s}{2s+d+1}}\right),
\end{aligned}$$

using the definition of j_0 and m_0. This completes the proof of the theorem. □

7 Numerical experiments

We now illustrate the usefulness of the adaptive nonlinear wavelet estimator described in Section 6.3 with the help of simulated and real-data examples. The overall numerical study presented below has been carried out in the `Matlab 7.7.0` programming environment.

7.1 Simulated data

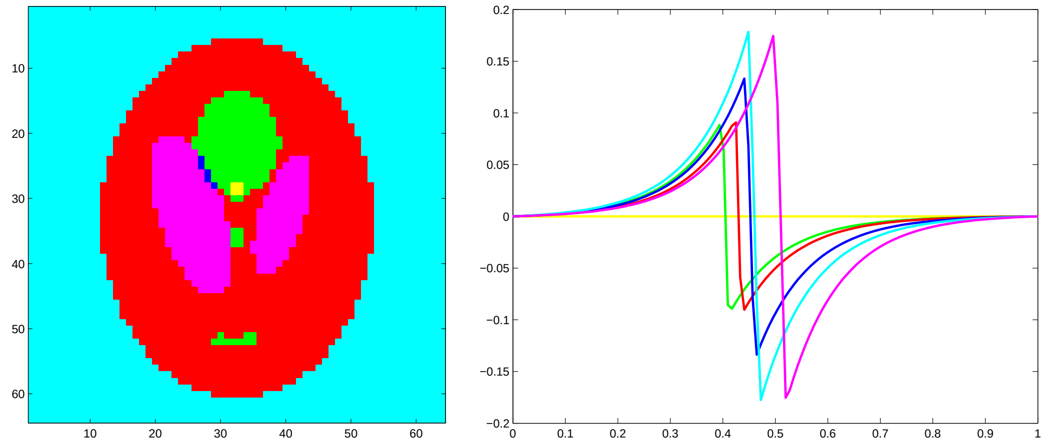
We have used as a synthetic 2-dimensional (2D) example the Shepp-Logan phantom image (see [Jain, 1989]) of size $N \times N$, with $N = 64$ displayed in Figure 7.3(a). This image is made of piecewise constant regions with different shape that partition the $N \times N$ pixels into 6 regions represented by different colors in Figure 7.3(a). To each pixel of a given region, we associate a one-dimensional (1D) signal of length $n = 128$. In this way, we are able to create a time-dependent 2D image $\left(\boldsymbol{f}(t_\ell, \underline{x}_{(k_1,k_2)})\right)_{1 \le \ell \le n, 1 \le k_1,k_2 \le N}$ that can be considered as the discretization of a function $\boldsymbol{f} : [0,1] \times [0,1]^2 \to \mathbb{R}$, with $t_\ell = \frac{\ell}{n}$ and $\underline{x}_{(k_1,k_2)} = \left(\frac{k_1}{n}, \frac{k_2}{n}\right)$.

Then, we have created noisy data from the model

$$Y_{\ell,(k_1,k_2)} = \boldsymbol{f}(t_\ell, \underline{x}_{(k_1,k_2)}) + \sigma w_{\ell,(k_1,k_2)}, \; 1 \le \ell \le n, \; 1 \le k_1, k_2 \le N, \tag{7.1}$$

where the $w_{\ell,(k_1,k_2)}$'s are i.i.d. standard Gaussian random variables, and $\sigma^2 > 0$ is the variance in the measurements ranging from a low to a high level in the simulations (we took signal-to-noise ratios equal to 7, 5 and 3). It is well known in nonparametric statistics (see e.g. [Brown and Low, 1996]) that there exists an asymptotic equivalence (in Le Cam sense) between the regression model (7.1) on nN^2 equi-spaced points, for each fixed $t \in T$, and the white noise model (1.1), when taking $\epsilon = \frac{\sigma}{\sqrt{nN^2}}$. Therefore, thanks to this asymptotic equivalence, one can use the 2D+time dependent wavelet block thresholding approach described in Section 6.3 to denoise data from model (7.1). To show the benefits of our approach, we compare it to two other mehods:

- pixel by pixel denoising based on 1D wavelet thresholding: for each fixed pixel (k_1, k_2), we apply a standard 1D wavelet-based denoising procedure with the universal threshold to the 1D data $\left(Y_{\ell,(k_1,k_2)}\right)_{1 \le \ell \le n}$,



(a) (b)

Figure 7.3: Synthetic example: (a) Shepp-Logan phantom made of 6 differents regions; (b) 1D signals associated to each region of the phantom image.

- slice by slice denoising based on 2D wavelet thresholding: for each fixed time t_ℓ, we apply a standard 2D wavelet-based denoising procedure with the universal threshold to the 2D data $\left(Y_{\ell,(k_1,k_2)}\right)_{1\leq k_1,k_2\leq N}$.

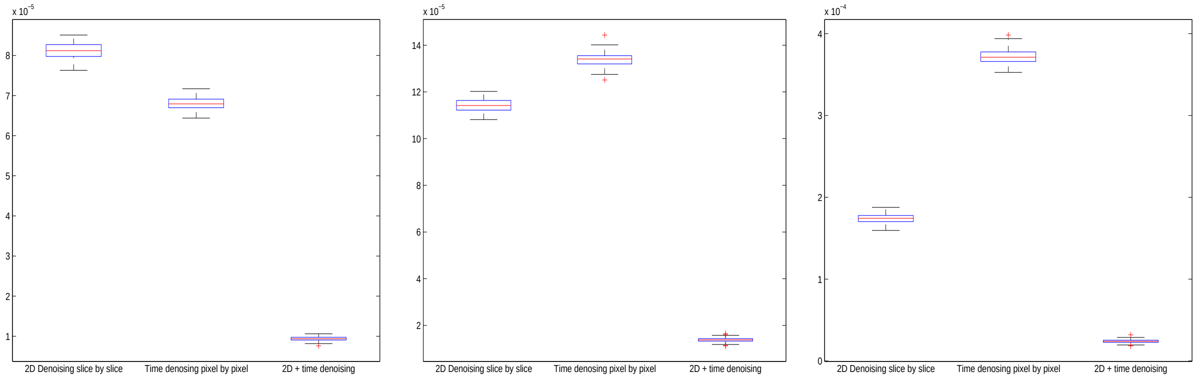
Then, we have generated $M = 100$ repetitions of model (7.1) for three different values of the considered signal-to-noise ratio. For each replication, the quality of an estimate $\hat{\boldsymbol{f}}$ obtained by one of the above described methods is measured via its empirical mean squared error

$$\frac{1}{nN^2}\sum_{\ell=1}^{n}\sum_{k_1,k_2=1}^{N}\left(\hat{\boldsymbol{f}}(t_\ell,\underline{x}_{(k_1,k_2)})-\boldsymbol{f}(t_\ell,\underline{x}_{(k_1,k_2)})\right)^2. \tag{7.2}$$

The results of these simulations are displayed in Figure 7.1 in the form of boxplots of the empirical mean squared error. Clearly, our approach yields the best results. The benefits of our method can also be clearly seen from the images displayed in Figure 7.5 which show temporal cuts of the various estimators for a given simulation of the model.

7.2 Real data

Now, we return to the real-data example on satellite remote sensing data discussed in Section 2. To apply the suggested adaptive nonlinear wavelet estimator, it is necessary to estimate the level of noise in the measurements. For this purpose, we estimate the level of noise in each 2D image at each wavelength using the median absolute deviation (MAD) of the empirical 2D wavelet coefficients at the highest level of resolution (see [Antoniadis et al., 2001] for further details on this procedure). Then, to apply our method, we took $\epsilon = \frac{\hat{\sigma}}{\sqrt{nN^2}}$ with $\hat{\sigma}$ being the maximum of these estimated values by MAD over the $n = 128$ wavelength, with $N = 64$. The result of our denoising procedure is displayed in Figure 7.6.



(a) (b) (c)

Figure 7.4: Boxplot of the empirical mean squared (7.2) error over $M = 100$ simulations from model (7.1) for the three methods (from left to right: pixel by pixel denoising based on 1D wavelet thresholding, slice by slice denoising based on 2D wavelet thresholding, 2D + time wavelet block thresholding) and for various values of the signal-to-noise ratio (SNR): (a) $SNR = 7$; (b) $SNR = 5$; (c) $SNR = 3$.

8 Concluding remarks

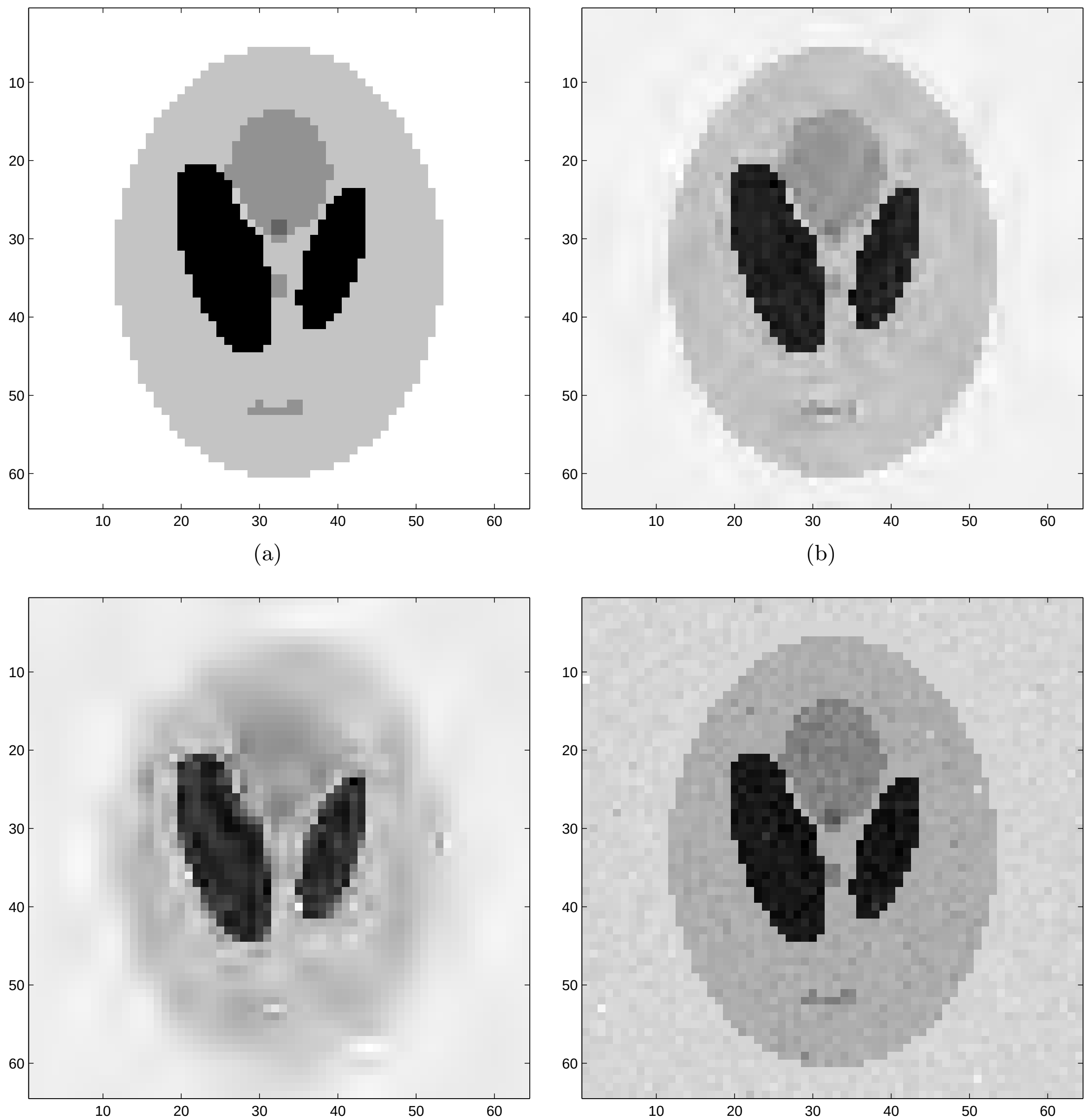
We considered the nonparametric estimation problem of time-dependent multivariate functions observed in a presence of additive cylindrical Gaussian white noise of a small intensity. We derived minimax lower bounds for the L^2-risk in the proposed spatio-temporal model as the intensity goes to zero, when the underlying unknown response function is assumed to belong to a ball of appropriately constructed inhomogeneous time-dependent multivariate functions. The choice of this class of functions was motivated by real-data examples and illustrated with the help of an example on satellite remote sensing data. We also proposed both non-adaptive linear and adaptive non-linear wavelet estimators that are asymptotically optimal (in the minimax sense) in a wide range of the so-constructed balls of inhomogeneous time-dependent multivariate functions. The usefulness of the suggested adaptive nonlinear wavelet estimator was illustrated with the help of simulated and real-data examples.

Some extensions of the present work are possible. They are briefly mentioned below.

[`Inverse Problems`] Model (1.1) can be extended to the case where the signal is observed through a linear operator plus noise. More precisely, one can consider the nonparametric estimation problem of time-dependent multivariate functions observed through a known or unknown linear operator with kernel $k(\underline{x}, \underline{u})$ and in a presence of additive cylindrical Gaussian white noise, namely

$$dY_\epsilon(t, \underline{x}) = \left(\int_{\mathcal{X}} k(\underline{x}, \underline{u}) \boldsymbol{f}(t, \underline{u}) d\underline{u} \right) d\underline{x} + \epsilon dW(t, \underline{x}), \tag{8.1}$$

where, as earlier, $t \in T$ (T is a compact subset of $\mathbb{R}$) is the time variable, $\underline{x} \in \mathcal{X}$ ($\mathcal{X}$ is a compact subset of $\mathbb{R}^d$, $d \geq 1$) is the space variable, $\boldsymbol{f} \in \mathbb{L}^2(T \times \mathcal{X})$ is the time-dependent multivariate function that we wish to estimate, $dW(t, \underline{x})$ is a cylindrical orthogonal Gaussian random measure (representing additive noise in the measurements), and $\epsilon > 0$ is a small level of noise, that may



(a) (b)

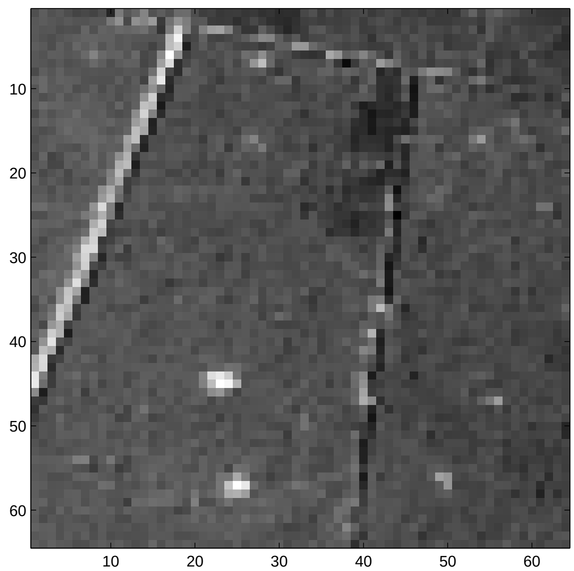
(c) (d)

Figure 7.5: Typical estimates obtained by the three methods with a signal-to-noise ratio $SNR = 5$ for a temporal cut at $t_\ell = 0.5748$: (a) true image without additive noise; (b) 2D + time wavelet block thresholding (our method); (c) slice by slice denoising based on 2D wavelet thresholding; (d) pixel by pixel denoising based on 1D wavelet thresholding.

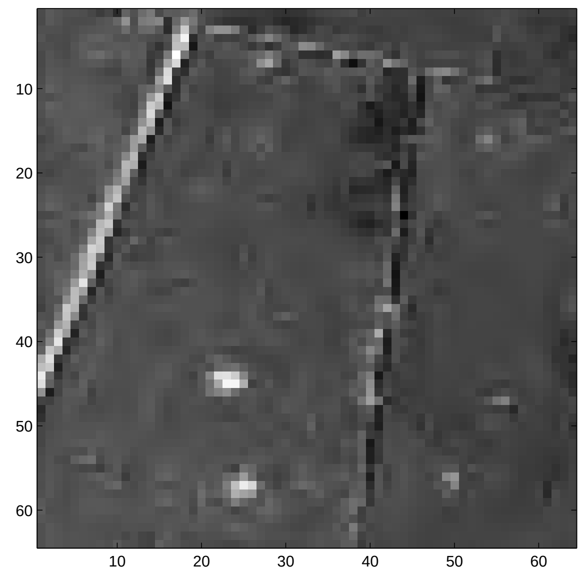
let be going to zero for studying asymptotic properties. An important example of kernel is the case where

$$k(\underline{x}, \underline{u}) = h(\underline{u} - \underline{x}) \quad \text{for some function} \quad h : \mathbb{R}^2 \to \mathbb{R},$$

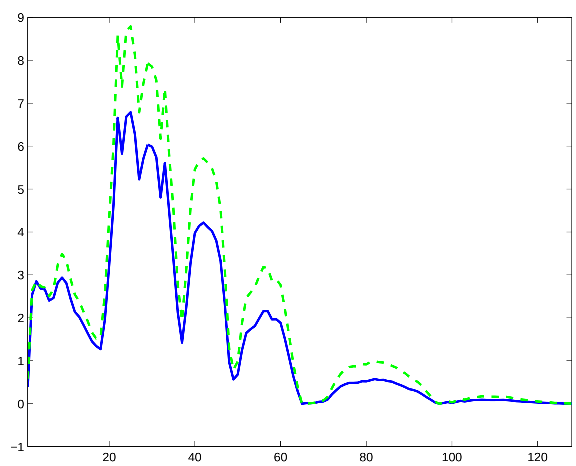
(with known or unknown singular values) leading to a time-dependent multivariate deconvolution



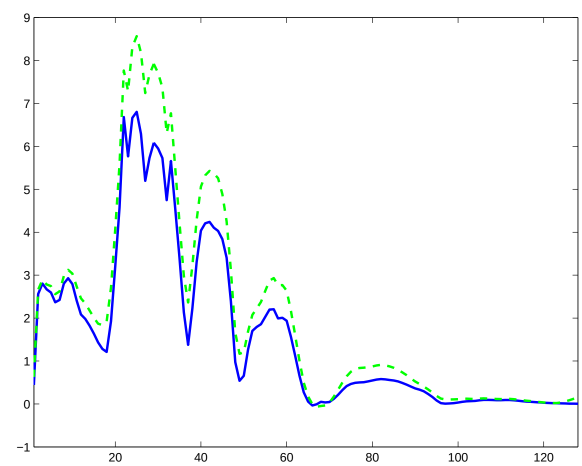
(a)



(b)



(c)



(d)

Figure 7.6: Satellite remote sensing image (64×64 pixels, over 128 wavelengths): (a) a 2D image measured at a specific wavelength (raw data); (b) 2D image at the same wavelength obtained after applying our method; (c) evolution over wavelength of the intensities of the two pixels in green and blue shown in Figure 2.1 (raw data); (d) intensities of these two pixels after denoising by our method.

problem. (Note that a sub-class of this model is the case of direct noisy observations of the time-dependent multivariate functions $\boldsymbol{f}(t, \underline{x})$, $t \in T$, $\underline{x} \in \mathcal{X}$, namely model (1.1) considered in this work.)

[`Smoothness Assumption`] In either model (1.1) or model (8.1), instead of using the standard (isotropic) d-dimensional Besov spaces on $\mathcal{X}$ to describe the smoothness of the underlying unknown response function $\boldsymbol{f}(t, \underline{x})$, for each fixed $t \in T$, one could consider anisotropic d-

dimensional Besov spaces on $\mathcal{X}$, where different smoothness is assumed in each direction (see, e.g., [Kyriazis, 2004]). Another possibility, is to consider subclasses of the so-called decomposition spaces that cover both the cases of standard (isotropic) d-dimensional Besov spaces as well as, in the case when $d = 2$, smoothness spaces corresponding to curvelet-type constructions (see [Chesneau et al., 2010]).

The above extensions are projects for future work that we hope to address elsewhere.

9 Appendix

9.1 Besov space and wavelet approximations

Lemma 1. *Let $A_1 > 0$ and $A_2 > 0$ be constants. Let $s_1 > 0$ and $s_2 > 0$ be the smoothness parameters in the space and time domains, respectively, such that $0 < s_1 < \tau_1$ and $0 < s_2 < \tau_2$, where τ_1 and τ_2 are the regularity parameters of the wavelet systems (ϕ, ψ) and $(\tilde{\phi}, \tilde{\psi})$, respectively. Let $1 \leq p_1, q_1 \leq +\infty$, $1 \leq p_2, q_2 \leq +\infty$. Assume that $\boldsymbol{f} \in \mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)$. Define $s'_1 = s_1 + d(1/2 - 1/p'_1) > 0$ with $p'_1 = \min(p_1, 2)$ and $s'_2 = s_2 + 1/2 - 1/p'_2 > 0$ with $p'_2 = \min(p_2, 2)$. Let $\boldsymbol{\alpha}_\lambda(t)$ be defined as in (4.1), and $\tilde{\alpha}_{\lambda,m,\ell}$ be defined as in (4.2). Then, for every $-1 \leq j < +\infty$,*

$$\sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2 \leq K'_1 A_1^2 2^{-2js'_1},$$

for some constant $K'_1 > 0$, only depending on s_1 and p_1, and for every $-1 \leq m < +\infty$,

$$\sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \leq K'_2 A_\lambda^2 2^{-2ms'_2},$$

for some constant $K'_2 > 0$, only depending on s_2 and p_2.

Proof. Since, for each $t \in T$, $\boldsymbol{f}(t, \cdot) \in B^{s_1}_{p_1,q_1}(A_1)$ with $1 \leq p_1, q_1 \leq +\infty$, using standard embedding properties of Besov spaces, there exists a constant $K'_1 > 0$, only depending on s_1 and p_1, such that for every $-1 \leq j < +\infty$

$$\sum_{|\lambda|=j} |\boldsymbol{\alpha}_\lambda(t)|^2 \leq K'_1 A_1^2 2^{-2js'_1}.$$

By the definition of $\mathbf{B}^{s_1,s_2}_{p,q}(A_1, A_2)$, and using standard embedding properties of Besov spaces, there exists a constant $K'_2 > 0$, only depending on s_2 and p_2, such that for every $-1 \leq m < +\infty$

$$\sum_{\ell=0}^{2^m-1} |\tilde{\alpha}_{\lambda,m,\ell}|^2 \leq K'_2 A_\lambda^2 2^{-2ms'_2}.$$

This completes the proof of the lemma. ∎

9.2 A large deviation inequality

Lemma 2. *Let $\delta > 1$. Then, for any $-1 \leq m < +\infty$ and $r \in A_m$,*

$$\mathbb{P}\left(\sum_{\ell \in U_{mr}} |y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}|^2 \geq \delta^2 \epsilon^2 L_\epsilon \right) \leq \epsilon^{(\delta-1)^2}. \tag{9.1}$$

Proof. The proof is inspired by the arguments used in the proof of Lemma 2 in [Pensky and Sapatinas, 2009]. Consider the set of vectors

$$\Omega_{m,r} = \left\{ v_\ell \in \mathbb{R} \setminus \{0\} \ : \ \sum_{\ell \in U_{m,r}} v_\ell^2 \leq 1 \right\},$$

and the centered Gaussian process defined by

$$Z_{m,r}(v) = \sum_{\ell \in U_{m,r}} v_m (y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}).$$

By Lemma 5 in [Pensky and Sapatinas, 2009], we need to find upper bounds for $\mathbb{E}\left(\sup_{v \in \Omega_{m,r}} Z_{m,r}(v)\right)$ and $\sup_{v \in \Omega_{m,r}} \mathrm{Var}\left(Z_{m,r}(v)\right)$. By the Cauchy-Schwartz inequality,

$$\begin{aligned} \sup_{v \in \Omega_{m,r}} Z_{m,r}(v) &= \sup_{v \in \Omega_{m,r}} \sum_{\ell \in U_{m,r}} v_m (y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}) \\ &= \left(\sum_{\ell \in U_{m,r}} (y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell})^2 \right)^{1/2}. \end{aligned}$$

Furthermore, Jensen's inequality implies that

$$\begin{aligned} \mathbb{E}\left(\sup_{v \in \Omega_{m,r}} Z_{m,r}(v) \right) &= \mathbb{E}\left(\sum_{\ell \in U_{m,r}} (y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell})^2 \right)^{1/2} \\ &\leq \left(\sum_{\ell \in U_{m,r}} \mathbb{E}(y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell})^2 \right)^{1/2} = \epsilon L_\epsilon^{1/2}. \end{aligned}$$

By independence of $y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell}$ and $y_{\lambda,m,\ell'} - \tilde{\alpha}_{\lambda,m,\ell'}$ for $\ell \neq \ell'$, we obtain

$$\sup_{v \in \Omega_{m,r}} \mathrm{Var}\left(Z_{m,r}(v)\right) = \sup_{v \in \Omega_{m,r}} \sum_{\ell \in U_{m,r}} v_\ell^2 \mathrm{Var}(y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell})^2 \leq \epsilon^2.$$

Thus, by Lemma 5 in [Pensky and Sapatinas, 2009], one has that

$$\mathbb{P}\left(\left(\sum_{\ell \in U_{m,r}} (y_{\lambda,m,\ell} - \tilde{\alpha}_{\lambda,m,\ell})^2 \right)^{1/2} \geq x + \epsilon L_\epsilon^{1/2} \right) \leq \exp\left(-\frac{x^2}{2\epsilon^2} \right),$$

for any $x > 0$. Finally, by taking $x = (\delta - 1)\epsilon L_\epsilon^{1/2}$, we arrive at (9.1), thus completing the proof of the lemma. ∎

Acknowledgements

Jérémie Bigot is grateful for the hospitality of the Department of Mathematics and Statistics at the University of Cyprus, Cyprus, and Theofanis Sapatinas is grateful for the hospitality of DMIA/ISAE at Toulouse University, France, where parts of this work were carried out. The authors would like to acknowledge helpful discussions with Gérard Kerkyacharian, Marianna Pensky and Dominique Picard at an early stage of this work.

References

[Antoniadis et al., 2001] Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: A comparative simulation study. *Journal of Statistical Software*, 6(6):1–83.

[Antoniadis et al., 2009] Antoniadis, A., Bigot, J., and von Sachs, R. (2009). A multiscale approach for statistical characterization of functional images. *J. Comput. Graph. Statist.*, 18(1):216–237.

[Brown and Low, 1996] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398.

[Chesneau et al., 2010] Chesneau, C., Fadili, J., and Starck, J.-L. (2010). Stein block thresholding for image denoising. *Appl. Comput. Harmon. Anal.*, 28(1):67–88.

[Chow et al., 2001] Chow, P.-L., Khasminskii, R., and Liptser, R. (2001). On estimation of time dependent spatial signal in Gaussian white noise. *Stochastic Process. Appl.*, 96(1):161–175.

[Härdle et al., 1998] Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

[Jain, 1989] Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[Ju et al., 2005] Ju, J., Gopal, S., and Kolaczyk, E. (2005). On the choice of spatial and categorical scale in remote sensing land cover characterization. *Remote Sensing of Environment*, 96(1):62–77.

[Klemelä, 2009] Klemelä, J. (2009). *Smoothing of multivariate data*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, NJ. Density estimation and visualization.

[Korostelëv and Tsybakov, 1993] Korostelëv, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

[Kyriazis, 2004] Kyriazis, G. (2004). Multilevel characterizations of anisotropic function spaces. *SIAM J. Math. Anal.*, 36(2):441–462.

[Mallat, 2009] Mallat, S. (2009). *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.

[Ou et al., 2009] Ou, W., Hämäläinen, M., and Golland, P. (2009). A distributed spatio-temporal eeg/meg inverse solver. *Neuroimage*, 44(3):932–946.

[Pensky and Sapatinas, 2009] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic setting: uniform case. *Ann. Statist.*, 37(1):73–104.

[Tsybakov, 2009] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[Wahba, 1990] Wahba, G. (1990). *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[Whitcher et al., 2005] Whitcher, B., Schwarz, A. J., Barjat, H., Smart, S. C., Grundy, R. I., and James, M. F. (2005). Wavelet-based cluster analysis: Data-driven grouping of voxel time-courses with application to perfusion-weighted and pharmacological mri of the rat brain. *Neuroimage*, 24(2):281–295.